

Modeling Heterogeneity in Indirect Effects: Multilevel Structural Equation Modeling Strategies

By

Emily Fall

Submitted to the Psychology and the
Faculty of the Graduate School of the University of Kansas
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Kristopher J. Preacher, Chairperson

Todd D. Little

Committee members

Pascal R. Deboeck

Wei Wu

Paul Johnson

Date defended: _____

The Dissertation Committee for Emily Fall certifies
that this is the approved version of the following dissertation :

Modeling Heterogeneity in Indirect Effects: Multilevel Structural Equation Modeling Strategies

Kristopher J. Preacher, Chairperson

Date approved: _____

Abstract

The heterogeneity implicit in much of social science research can be accommodated by using complex modeling procedures such as SEM or MLM. Ignoring heterogeneity, particularly with regard to nested data structures, can have serious consequences for model estimation and lead to incorrect conclusions about tested hypotheses. In mediation models, the consequences of ignoring nesting can have a substantial impact on the indirect effect. Inflated standard errors and bias in the parameter estimates lead to inaccurate estimates of the indirect effect, as well as reduced power to detect the effect. Using multilevel structural equation modeling (MSEM), data were generated based on a cross-lagged panel model for mediation. By fitting a single-level model to the data, the consequences for the estimation and detection of the indirect effect when heterogeneity is ignored is examined through measures of relative bias, power, and model fit.

Acknowledgements

My sincerest thanks go to my committee for bearing with me through this whole process.

To my advisor, Dr. Kristopher Preacher, without whose guidance and unfailing support I could not have succeeded. While an advisor is essential to any academic endeavor, in my case, clearly not just any advisor would do. Thanks for being the one who finally got me through, for the countless drafts you read last minute, and for the patience and encouragement that followed every moment of panic I had.

To Dr. Todd Little, who not only held on and made sure there was a program for me to graduate from but worked to build one that I am exceedingly proud to have been a part of. Thanks for continuing to believe in both the program and me.

To Dr. Pascal Deboeck and Dr. Wei Wu, who graciously agreed serve on my committee and whose mentorship contributed immeasurably to my success. Thanks for coming to KU and taking a chance on all of us. I am so grateful that I stayed long enough to learn from both of you.

To Dr. Paul Johnson, without whom I would still be formatting this document and who challenges me not to accept common standards or take the easy way out. Thanks for stepping in when I needed it. Enrolling in that political science course was one of the best decisions I made at KU.

Special thanks to my husband, Spencer Cramer, who endured the whole process with incredible patience and understanding, who always made sure that I had food to eat, and who bore my stress as if it were his own. I cannot imagine getting through this

without you.

Finally, though there are too many to mention, an enormous thank you to all of the educators, friends, and family who have inspired, challenged, and supported me. Immanuel Kant said, “Science is organized knowledge. Wisdom is organized life.” Whatever progress I’ve made toward either is a direct result of the influence of my mentors in my education and in my life.

Contents

1	Introduction	10
2	Review of Literature	13
2.1	Methods for Assessing Mediation	13
2.1.1	Causal Steps	13
2.1.2	Recent Advances	14
2.1.3	Mediation in SEM	17
2.1.4	Mediation in MLM	19
2.2	The Importance of Longitudinal Data	21
2.2.1	Longitudinal Data and Causality	21
2.2.2	Heterogeneity in Longitudinal Data	22
2.2.3	Heterogeneity in Multilevel Models	23
2.2.4	Models for Longitudinal Mediation	25
2.3	Multilevel Structural Equation Modeling	28
2.3.1	Foundations and Development	28
2.3.2	Goldstein and McDonald (1986, 1988)	29
2.3.3	Muthén’s (1989) MUML Method	31
2.3.4	Current Models and Estimation	33
2.3.5	Other issues related to MSEM	35
2.3.6	Mediation in MSEM	38

2.3.7	Current Study	40
3	Methods	42
3.1	Simulation Parameters	42
3.1.1	Determining Bias	44
3.1.2	Confidence Interval Coverage and Power	45
3.1.3	Model Fit	45
3.2	Results	45
3.2.1	Bias	45
3.2.2	Confidence Interval Coverage and Power	47
3.2.3	Model Fit	48
3.3	Discussion	49
3.4	Future Directions	50
3.5	Conclusion	51

List of Figures

1.1	Mediation Model	11
2.1	Cross-lagged Panel Model	28
2.2	An example of an MSEM model.	36
3.1	MSEM Simulation Data Generating Model	43

List of Tables

3.1	Bias in the Indirect Effect	46
3.2	Bias in the Estimated Standard Error of the Indirect Effect	47
3.3	Coverage and Power	48
3.4	Model Fit	49

Chapter 1

Introduction

The business of statistics is heterogeneity. From very early in an introductory statistics course, the pursuit begins with a z -test comparing the difference between some sample value and a population value, and progresses to differences between two samples, either between groups or in repeated measures on individuals. Simply finding differences, however, is not enough. Statistical methods like regression, analysis of variance, and all the complex modeling procedures that follow in the general linear modeling family are designed to capture not only whether or not there are differences, but also how big those differences are and what causes the differences.

Heterogeneity itself is heterogeneous. Some of the types of heterogeneity that are of interest to researchers include differences between individuals at a single time point, differences within individuals over time, differences in mean levels of a trait, and differences in variability. Often several types of heterogeneity are investigated simultaneously. In the current study, I propose to model heterogeneity in indirect effects within and between individuals and groups across time using multilevel structural equation modeling (MSEM). Prior research providing the basis for this study is outlined in three major sections: a review of methods for assessing mediation, the importance of longitudinal data for examining heterogeneity and causality, and the development of and current methods for MSEM.

The first section that follows is a review of methods for assessing mediation. The target of

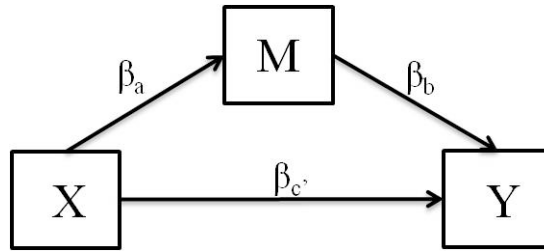


Figure 1.1: Mediation Model

much of the research in the behavioral sciences involves complex processes and intervening variables, which makes mediation analysis a useful tool in many modeling procedures. As a result, a large amount of work has been conducted to determine best practices for mediation. Baron and Kenny (1986) introduced the idea of mediation as distinct from moderation, and emphasized that mediation was a causal process by which a predictor variable, X , influences or explains variability in an outcome variable, Y , by way of a mediator variable, M (see Figure 1.1).

Since 1986, Baron and Kenny's work has been cited over 23,000 times, and has been extended from the simple, single-level regression model they proposed into multilevel modeling and structural equation modeling frameworks. The implications for inference and understanding causality as well as advantages and limitations for handling heterogeneity in SEM and MLM are also addressed in this section of the literature review.

In the second section, issues related to the measurement of change over time, the importance of longitudinal data for understanding causality, and the types of heterogeneity that can be modeled using longitudinal data are addressed. Establishing causality necessarily requires the passage of time, whether it is a matter of milliseconds, as a person reacting to a sudden sound, or years, as with many developmental outcomes. Mediation assumes at least two causal paths, shown in Figure 1, and time becomes an even more important consideration in the specification and interpretation of these models. In addition to the benefits of accurately framing questions of causality, longitudinal data also provide a variety of sources of variability, including variability across time and across person, which must be addressed. Therefore, the methods proposed in this study deal specifically with longitudinal data.

Finally, although MLM and SEM provide distinct advantages for assessing mediation in longitudinal data, limitations exist. However, some of the limitations of each approach can be overcome by multilevel structural equation modeling (MSEM). The development of and current methods for MSEM are presented in section three. Then methods for assessing mediation in MSEM developed by Preacher, Zyphur, and Zhang (2010) are discussed. These methods, while at the forefront of research on both MSEM and mediation, have not been extended to longitudinal data. It is for this reason that an alternative method is proposed in the current study for modeling heterogeneity in indirect effects for longitudinal data in MSEM.

Chapter 2

Review of Literature

2.1 Methods for Assessing Mediation

2.1.1 Causal Steps

Baron and Kenny (1986) defined mediation in the context of psychological research as a process by which a third variable accounts for the relationship between a predictor and an outcome variable. The authors outline three conditions that must be met in order for a variable to function as a mediator. First, the independent variable is a significant predictor of the mediator. Second, the mediator is a significant predictor of the dependent variable, when controlling for the independent variable. Finally, the inclusion of the mediator in the model changes the significant relationship between the independent and dependent variables such that it is no longer significant (Baron & Kenny, 1986). If these conditions hold, a test of the indirect effect can then be performed using a test of significance of the product term $\beta_a\beta_b$, where β_a is the regression coefficient of M on X and β_b is the regression coefficient of Y on M , using the standard error calculation proposed by Sobel (1982), which is given as:

$$\sigma_{ab} = \sqrt{a^2\sigma_b^2 + b^2\sigma_a^2}. \quad (2.1)$$

This causal steps approach has been one of the most widely used methods for testing mediation but is highly criticized in recent literature for a lack of power to detect the indirect effect, and relying on implied relationships between variables rather than strict quantification of the indirect effect (Hayes, 2009). The low power of the causal steps approach is a result of the multiple significance tests required to detect the indirect effect. Additionally, models with multiple mediators or more than one mediated path require even more significance tests, which results in even lower power, and higher Type II error rates. Finally, while the Sobel test is commonly used and does provide a test of the indirect effect in $\beta_a\beta_b$, conditioning the use of the Sobel test on Baron and Kenny's criteria provides little additional information about the presence of mediation, making the Baron and Kenny criteria a less appropriate choice for tests of the indirect effect than some more recent advances (Hayes, 2009). For these reasons, recent work in mediation analysis is recommending other methods of calculating and testing the indirect effect which are now preferred over Baron and Kenny's causal steps approach.

MacKinnon et al. (2002) propose a joint significance test of only paths a and b , and found that compared to the causal steps approach of Baron and Kenny, and a similar approach proposed by Judd and Kenny (1981), the joint significance test provided the best balance of power and Type I error rates. However, this joint significance test suffers much the same criticism as other causal steps approaches. Also, the Sobel test assumes a normal sampling distribution, but is applied to a product term which tends to have an asymmetric distribution, which could lead to spurious results (Hayes, 2009; Pituch, Whittaker, & Stapleton, 2005).

2.1.2 Recent Advances

For simple, single-level regression models, several alternatives to Baron and Kenny's causal steps approach have been proposed. These methods can be grouped into two types of tests: difference in coefficients and product of coefficients.

Both the difference in coefficients and the product of coefficients approaches provide a point estimate of the indirect effect, and both methods involve estimating two separate equations. In

the difference of coefficients method, an equation predicting the outcome, Y , from the predictor variable, X , is estimated.

$$Y_i = \beta_{0yx} + \beta_c X_i + r_i \quad (2.2)$$

Then, an equation predicting the outcome, Y , from the both predictor variable, X , and the mediator variable, M , is estimated.

$$Y_i = \beta_{0mx} + \beta_{c'} X_i + \beta_b M_i + r_i \quad (2.3)$$

The difference in the regression coefficients, $\beta_c - \beta_{c'}$, for the predictor variable is the estimated indirect effect (Judd & Kenny, 1981).

In the second method a product of coefficients of two equations is calculated. First, an equation estimating the effect of X on M is fit.

$$M_i = \beta_{0mx} + \beta_a X_i + r_i \quad (2.4)$$

Next, as before, Equation 2.3, predicting Y from both X and M , is estimated. Then, the product of the coefficients, $\beta_a \beta_b$, is calculated as the point estimate of the indirect effect. These difference in coefficients and product of coefficients methods of point estimation for the indirect effect are algebraically equivalent in single-level models (MacKinnon et al., 1995).

The indirect effect, whether a difference or a product, can be tested in a number of ways, most of which involve calculating confidence intervals. Confidence intervals for products of coefficients methods can be obtained used the Sobel standard error; however the nonnormality of a distribution of a product makes the Sobel standard error a poor choice.

Alternatively, an asymmetric confidence limit approach has been proposed (MacKinnon et al., 2002). In this approach, z -scores are obtained for a and b . Then, the product $z_a z_b$ is calculated, and the asymmetric confidence interval is constructed based on upper and lower critical values obtained from the distribution of a product term of normally distributed variables rather than a

z - or t -distribution (Meeker et al., 1981; Pituch et al., 2005). By using asymmetric confidence intervals, the limitations of the normality assumption made by the Sobel test can be overcome; however, obtaining critical values for the distribution of the product is difficult without the use of tables, and reliance on tables means being limited to the increments used in established tables such as Meeker et al. (1981), which may require rounding or approximating to find the appropriate values (Hayes, 2009; Pituch et al., 2005).

One way to avoid using established tables or cumbersome methods for finding critical values to create confidence intervals is to use resampling techniques such as bootstrapping. Several bootstrapping methods, including the percentile method, the bias corrected method, and the accelerated bias corrected method, can be implemented to obtain more or less accurate confidence intervals (MacKinnon et al., 2004).

Bootstrapping involves resampling with replacement to generate a sampling distribution. From a sample, a single observation is drawn, recorded, and replaced. This process is repeated until the bootstrap sample reaches the same number of observations as the original sample. Because sampling with replacement is used, a single observation from the original sample may occur in the bootstrap sample more than once or not at all. Thousands of bootstrap samples are created to generate a sampling distribution of the desired statistic. From the sampling distribution, confidence intervals can be calculated in a number of ways, including a straightforward percentile method as well as methods based on z - or t -distributions. Bootstrapping works well in single-level models including structural equation models (Cheung & Lau, 2008), and as it has been shown to be more powerful and valid than many other methods, it is arguably the best method (Hayes, 2009; MacKinnon et al., 2004). One limitation of resampling techniques is that they become computationally prohibitive when models become increasingly complex or with clustered data, as in the case of multilevel models.

In order to generate sampling distributions in situations where bootstrapping and other resampling methods are not feasible, a Monte Carlo method has been proposed by MacKinnon et al. (2004). The Monte Carlo method for assessing mediation (MCMAM), involves three steps. First,

β_a and β_b and their standard errors are estimated. From the sample estimates a sampling distribution consisting of a large number random samples of the product of β_a and β_b , with population values equal to the sample values β_a , β_b , σ_{β_a} , and σ_{β_b} , is obtained. MacKinnon et al. (2004) suggest 1000 random samples, but with modern computing methods in programs such as Mplus (Muthén & Muthén, 2010) and R (R Development Core Team, 2008), 5000 or even 10,000 random samples are equally easy to generate. The confidence interval is then calculated based on the values in the sampling distribution using percentiles for upper and lower confidence limits (MacKinnon et al., 2004).

Each of these methods has advantages and disadvantages in the single-level model, and perform well, to varying degrees, when applied to structural equation modeling and multilevel modeling (Krull & MacKinnon, 1999; MacKinnon et al., 2002; Pituch et al., 2005).

2.1.3 Mediation in SEM

Mediation models use path analysis to test a number of simultaneous equations between three or more variables. Each of these variables could easily be measured by multiple indicator variables to reduce measurement error. Thus, a simple path analysis becomes a latent variable path analysis, making SEM a logical framework for assessing mediation.

The utility of SEM has led to its widespread use across a number of disciplines and research settings. Models in the SEM framework can be either exploratory or confirmatory, include both manifest and latent variables, and allow for the analysis of means as well as covariance structures, multivariate outcomes, nonlinear relationships, path models, and many other complex relationships. SEM also has many well accepted and understood measures of model fit. Like MLM, SEM can also be applied to a number of experimental designs, including both standard cross-sectional and longitudinal data structures.

For each SEM analysis, two submodels are commonly estimated. The measurement model, which defines the relationships between measured and latent variables, is commonly expressed as:

$$y = \nu + \Lambda\eta + \varepsilon, \quad (2.5)$$

where y is a vector of observed variables, ν is a vector of measurement intercepts, Λ is a matrix of factor loadings of measured variables on latent factors, η is a matrix of factor scores for the latent variables, and ε is a vector of residuals, with a covariance matrix Θ (Bovaird, 2007; Curran, 2003; Jöreskog & Sörbom, 1993).

The structural model, then, is the portion of the model that defines the relationships among the latent factors in the model, and is given as:

$$\eta = \mu + \beta\eta + \zeta, \quad (2.6)$$

where μ is a vector of means for the latent factors, β is a matrix of regression coefficients for the latent factors, and ζ is a vector of latent residuals, with covariance matrix Ψ . The structural model can be used to test any number of simultaneous equations or complex relationships, including single-level regression models, path analysis with mediation, higher-order latent factors predicting first-order latent variables, and growth curve models, to name a few.

For testing mediation in SEM, Cole and Maxwell (2003) recommend five general steps, the first of which is simply fitting a measurement model. The next three steps involve testing the structure of the model for factorial invariance, omitted variables, and omitted paths. These steps are important for determining if the relationships between measured and latent variables are consistent over time or across groups, and that a sufficient number of variables and paths have been included in the model to adequately account for the variability in the model. Finally, a model that is tested and found to be complete can be used to assess mediation, and direct and indirect effects can be calculated and tested.

Bootstrapping has been demonstrated as an appropriate and efficient method for testing indirect

effects in SEM (Cheung & Lau, 2008). Although many researchers test for mediation by testing two separate paths for significance, from the independent variable to the mediator (β_a) and from the mediator to the dependent variable (β_b), this test alone does not provide sufficient support for an indirect effect. Therefore, Cheung and Lau (2008) recommend testing the significance of the indirect effect ($\beta_a\beta_b$). The authors also found that the bootstrap percentile confidence intervals had higher power and more accurate Type I error rates than methods based on Sobel tests and other standard error calculations. Therefore, bootstrapping for tests of indirect effects is recommended.

2.1.4 Mediation in MLM

Whereas SEM is useful for assessing mediation in a small number of groups by using models such as mean and covariance structures (MACS) models, it is limited when the number of groups is large and the dependence between observations is not ignorable. In this case, multilevel modeling (MLM) becomes the more advantageous approach. MLM, or hierarchical linear modeling, accounts for dependency in the data by including random components to model the correlations in the data (Roberts, 2004).

The level of dependency among the observations can be measured by the intraclass correlation (ICC). The ICC is the proportion of the overall variance in the dependent variable that is accounted for by clustering. Conceptually, the ICC is the degree to which observations in the same cluster are more similar to each other than they are to observations in another cluster. If the ICC is sufficiently large, the dependency in the data warrants concern over using standard OLS regression models (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999).

Multilevel models are specified by equations at two or more levels of data. Level-1 equations include random components and higher level equations include both fixed and random components at each level of the model. The standard level-1 equation for the MLM looks very similar to single-level regression and is given as:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}, \quad (2.7)$$

where i represents the level-1 unit, j represents the level-2 unit, β_{0j} represents an estimate of the intercept, β_{1j} represents the slope of y with respect to x , and e_{ij} is the level-1 residual term. The level-2 equations are given as:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad (2.8)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad (2.9)$$

where γ_{00} is the mean intercept and u_{0j} is the deviation from the intercept for the level-2 unit; likewise, γ_{10} is the mean slope and u_{1j} is the deviation from the slope for the level-1 unit. Including these terms allows for variances and covariances of random components to be estimated. The model can also be written as a single, expanded equation:

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + u_{1j}x_{ij} + e_{ij}, \quad (2.10)$$

These basic equations can be expanded easily to include more predictors and more levels of nesting to accommodate many different types of research questions and sampling frameworks. The most common applications are in two- and three-level cross-sectional designs and longitudinal designs.

A key difference between single- and multilevel models with respect to mediation analysis is that MLM allows for paths between only level-1 variables or between a level-2 independent variable and level-1 mediator and dependent variable. Each variable in the path may affect another variable measured at the same level or at a lower level, but not at a higher level (Krull & MacKinnon, 1999). Krull and MacKinnon (1999) use 2-1-1 notation to illustrate the independent variable-mediator-dependent variable relationship in data designs. Krull and MacKinnon (2001) have argued that MLM is flexible enough to allow multiple mediation paths and multiple media-

tor variables in a single path, but the multiple mediator model is not easy to specify in the MLM framework (Preacher et al., 2010).

Despite the flexibility and power of MLM for mediation, some limitations exist. First, the between- and within-group effects are conflated, resulting in bias. Also, upper-level dependent variables cannot be accommodated (Preacher et al., 2010). Another consideration is that in MLM measurement error is not explicitly modeled. This can lead to underestimation of the effect of the independent variable on the dependent variable when the mediator is left out, or overestimation of the direct effect of the independent variable on the dependent variable (Cole & Maxwell, 2003).

Many of the methods for assessing mediation presented so far focus on cross-sectional data. These methods often misrepresent the causality implied by the model and can lead to biased estimates of the indirect effect. Change unfolds over time, so time must be addressed in any model that attempts to explain or predict change.

2.2 The Importance of Longitudinal Data

2.2.1 Longitudinal Data and Causality

Gollob and Reichardt (1987) lay out three principles for causal effects. First, causes need time to exert their effects. Simply put, causality cannot be assumed between two simultaneously occurring events. For example, a researcher would never assume that the differences in test scores between two groups are the result of an educational intervention implemented on the day of the testing. This first principle seems obvious enough, but it is commonly ignored as many cross-sectional studies make such inferences.

The second principle is that a variable can have an effect on itself at a later time (Gollob & Reichardt, 1987). This principle is included in regression models which use a pre-test score as a predictor of a post-test score. A variable is influenced by prior values of the same variable. Again, this seems a simple principle to illustrate. A person's IQ at age 12 will have some potential to predict his IQ at age 25. Therefore, prior measurements and autoregressive paths need to be

included in any model of causality.

The third principle is that the lag time between measurements influences effect size (Gollob & Reichardt, 1987). This principle is related to the logic behind the first principle. The magnitude of an effect can vary depending on the time that elapses between measurements. Therefore, measurements in longitudinal data should be spaced according to the expected amount of time it should take for a cause to have an effect.

The consideration of time is crucial for analyzing longitudinal data, but longitudinal data are essential for understanding causality. Cross-sectional models used to test causality have been shown to yield biased estimates of direct, indirect, and total effects in comparison to longitudinal autoregressive models (Gollob & Reichardt, 1987, 1991; Maxwell & Cole, 2007). The direction and magnitude of the bias in the parameter depends upon the specification of the model and the stability of the effects, and is largely the result of the lack of autoregressive effects in cross-sectional models. In addition to causing biased estimates, cross-sectional data used in mediation do not accurately represent the process of change implied by the model (Maxwell & Cole, 2007; Selig & Preacher, 2009). Further, some cross-sectional designs include time as a predictor of change, which implies that time itself is the reason for change rather than some intervention or mediating effect (Pitariu & Ployhart, 2010). For example, the change in a student's math ability from the beginning of the year to the end of the year is probably not merely a result of the passage of time. It is more likely the result of learning and practice that occurred in the time between measurements. In reality time is rarely the cause of change, but rather a metric to represent change. Therefore, longitudinal modeling is especially important for mediation.

2.2.2 Heterogeneity in Longitudinal Data

In addition to helping researchers properly address causality, longitudinal data provide a distinct advantage over cross-sectional data in the types of variability that can be modeled. Longitudinal data have two inherent types of heterogeneity: within-person across time, and between-person. When viewed as a multilevel model, characteristics of the level-2 unit, the individual, can have

a specific effect on change over time in level-1 variables. If these higher-level influences are not accounted for, then the within-person variability over time cannot be distinguished from the between-person variability found in a cross-sectional analysis (Hoffman & Stawski, 2009). MLM allows for both between- and within-person variability to be modeled across time.

SEM can also be used to test differences between groups, in at least three different ways: multi-group models, multilevel structural equation models, and mixture models. All of these methods account for heterogeneity in clustered data. MACS methods require a separate model to be fit for each group. Then models are tested for invariance across groups. If the number of groups is large, this method becomes difficult to estimate and a multilevel model may be preferred. However, multilevel models assume that the level-2 units, or groups, are randomly sampled from a sufficiently large population. If the number of available groups is small, this assumption may not be met, and the multigroup method would be preferred (Selig et al., 2008). In the case of longitudinal data, the level-2 units are persons rather than groups, so a population large enough from which to randomly sample level-2 units is less difficult to imagine, and a multilevel model would be more appropriate than a MACS model. Single-level SEM is not capable of modeling both within-person and between-person variability in a single model, and is therefore less appropriate for examining multiple sources of heterogeneity (Hoffman & Stawski, 2009).

2.2.3 Heterogeneity in Multilevel Models

Whether a study is designed to sample from clusters or not, the non-independence of data as the result of a grouping variable is a common data issue in the social sciences, and because units in the same group share common factors two types of heterogeneity must be considered. The within-level variance is the result of individual units within and across groups differing from each other by virtue of the fact that they are different units with individual traits and experiences. This within-level variance can be explained by variables that are specific to unit-level characteristics, such as gender, IQ, or time. The between-level variance is the result of individual units within a group differing, at mean level, from the units within another group. The between-level variance can be

explained by group-level variables such as teaching style in a classroom, family size, or, if person is the level-2 unit, math ability, or neuroticism. These two sources of variance contribute unique information about unit differences, and when one source is ignored the ability of a statistical model to explain what causes these differences is hindered.

The consequences of ignoring a level of clustering have been examined in a wide range of contexts. In the simple case of multilevel regression, ignoring a level and using ordinary least squares regression will result in small standard error estimates of the coefficients, which in turn inflates the Type I error rate for hypothesis testing (Hox, 2002). Additionally, variance parameters and some fixed effects can be overestimated and power is reduced (Moerbeek, 2004). These effects hold even in more complex models, but the more complex the model becomes, the more potential consequences arise. In addition to cross-sectional models for multilevel regression, Skinner and Toledo Vieira (2007) found that in longitudinal models where complex error structures can be specified, the effect of ignoring clustering on standard error estimation can be larger than for cross-sectional analyses. The consequences of ignoring a level of nesting in repeated measures data can be similar to those in cross-sectional data, but only if the errors are uncorrelated (Moerbeek, 2004).

Van den Noortgate, Opdenakker, and Onghena (2005) examined the effects of ignoring a level of nesting in models with three and four levels with balanced and unbalanced data. The authors found that for both the balanced and unbalanced cases, ignoring the highest level of clustering results in inflated variance estimates and larger standard errors for variance estimates at the level directly below the ignored level while other levels remain unaffected. Additionally, if an intermediate level is ignored, the variance estimates and associated standard errors at both the level above and the level below the ignored level are affected. The impact of ignoring a level on the fixed effects is lower; however, the standard errors of fixed effects in the level below an ignored level are increased, particularly for predictor variables. Moerbeek (2004) found similar effects noting that the magnitude of the effect on the standard error of predictor variables was not only dependent upon the ignored level, but also the level at which the predictor varies, and the values of the variance components themselves.

Some latent variable models such as growth mixture models (GMM) and multiple indicator multiple cause (MIMIC) models also have been examined in the multilevel context for the consequences of ignoring clustering (Chen et al., 2010; Finch & French, 2011). Chen, Kwok, Luo, and Willson (2010) found that, with GMM, ignoring a higher level of nesting resulted in less accurate classification and relative bias in the level-2 variance estimates. Similar to the findings of other research (Moerbeek, 2004; Van den Noortgate et al., 2005), the authors also found that ignoring clustering had little or no impact on estimates of fixed effects and a differential impact on the standard errors of the fixed effects dependent upon whether the fixed effect was an intercept or a slope.

2.2.4 Models for Longitudinal Mediation

MLM can easily accommodate longitudinal data. Repeated measurements on individuals are necessarily dependent, and just as students in the same classroom are likely to have more in common with each other than with students from another classroom, a person has more in common with him- or herself over time than with other individuals at any given time. Here, rather than individuals being level-1 units, each individual is a level-2 unit with the repeated measures on that individual as level-1 units. In this case, within- and between-group variance can be viewed as within- and between-person variance with longitudinal data.

The advantages of MLM for repeated measures are the same as for cross-sectional designs, and also include the ability to estimate complex error structures, which is also possible but not often justified in models for cross-sectional data (Snijders & Bosker, 1999). This flexibility allows the researcher to specify differing patterns of error variance estimation to account for changes in variance as the measurements of independent and dependent variables become more distant in time. Predictors in longitudinal MLM can be specified as either time-varying or time-invariant. Time-varying predictors are measured repeatedly across time and are level-1 predictors. Time-invariant predictors are those characteristics which do not change over time, such as gender or race, are measured only once, and are level-2 predictors (Hoffman & Stawski, 2009). Although MLM has

many analytical advantages, it also has some shortcomings. Using manifest variables inherently assumes that these predictors are free of measurement error, which is a difficult assumption to meet in reality. One correction for this is the use of multiple manifest variables as indicators of a latent construct as applied in structural equation modeling (Curran, 2003). However, the MLM framework does not easily allow for latent variables. Also, MLM is typically used only with univariate outcomes, whereas other methods such as SEM allow for multivariate outcomes (Mehta & Neale, 2005). Finally, MLM currently does not provide good overall measures of model fit (Curran, 2003).

SEM provides a flexible alternative to MLM for modeling mediation. Several types of models can be specified in SEM and can be used to assess mediation. One potential model for longitudinal mediation analysis is the latent growth mediation model (LGM; Cheong et al., 2003). The LGM can be used to model both inter- and intraindividual variability across time, and allows change over time to be a variable in the model as an independent, a dependent, or a mediator variable. The latent difference score (LDS) model is an alternative to the LGM approach. The LDS model allows change to vary across occasions (Selig & Preacher, 2009). In contrast to both the LGM and LDS approaches, the cross-lagged panel model (CLPM; Cole & Maxwell, 2003; Selig & Preacher, 2009) has distinct advantages for modeling interindividual differences across time. In this model time is not explicitly modeled, and the effect of lag between measurements can be costly if chosen poorly (Maxwell & Cole, 2007; Selig & Preacher, 2009).

The cross-lagged panel model (CLPM), developed largely by Campbell (1963) and presented by Cole and Maxwell (2003), is a common model for panel designs for longitudinal data. Multiple variables are measured at multiple time points, and a series of regression paths are used to predict a variable at time t from the same variable at time $t-1$ as well as from the other variables at earlier time points. The full CLPM is represented by the following three equations:

$$X_{[t]} = \beta_{X,[t-1]}X_{[t-1]} + \zeta_{X,[t]} \quad (2.11)$$

$$M_{[t]} = \beta_{M,[t-1]}M_{[t-1]} + \beta_{X,[t-1]}X_{[t-1]} + \zeta_{M,[t]} \quad (2.12)$$

$$Y_{[t]} = \beta_{Y1,[t-1]}Y_{[t-1]} + \beta_{M,[t-1]}M_{[t-1]} + \beta_{X,[t-2]}X_{[t-2]} + \zeta_{Y,[t]} \quad (2.13)$$

The CLPM, as a path model, can be fit in an SEM framework and is well suited for mediation models (see Figure 2.1). Three types of effects can be calculated from path models: direct effects, indirect effects, and total effects. Further, Gollob and Reichardt (1991) suggest that for longitudinal mediation models two types of indirect effects, time-specific and total indirect effects, can be calculated. In Figure 2, an example of a direct effect is X_1 on Y_3 and is estimated by (e.g.) β_{xy1} . A time-specific indirect effect is calculated with a single set of paths from one variable to another. For example, a time-specific indirect effect from X_1 to Y_4 can be calculated from the product of three path coefficients connecting the two variables: $\beta_{xm1} \times \beta_{my2} \times \beta_{y3}$. The total indirect effect is the sum of all of the time-specific indirect effects linking two variables. So from X_1 to Y_4 the total indirect effect is calculated from three paths: $X_1 \rightarrow M_2 \rightarrow Y_3 \rightarrow Y_4$; $X_1 \rightarrow M_2 \rightarrow M_3 \rightarrow Y_4$; and $X_1 \rightarrow X_2 \rightarrow M_3 \rightarrow Y_4$. Finally, the total effect can be calculated as the sum of the total indirect effect and the direct effect.

Both MLM and SEM have limitations for assessing mediation. Single-level SEM does not allow for the effect of higher-level clustering, which results in aggregation of the data and a loss of information. The problem of separating between- from within-person effects in SEM has received very little attention (Hoffman & Stawski, 2009). Whereas MLM allows for distinct between- and within-person effects to be modeled, it does not allow for a level-2 outcome variable to be modeled, and between- and within-effects can be conflated. Therefore, Preacher et al. (2010) recommend MSEM as a method for assessing mediation.

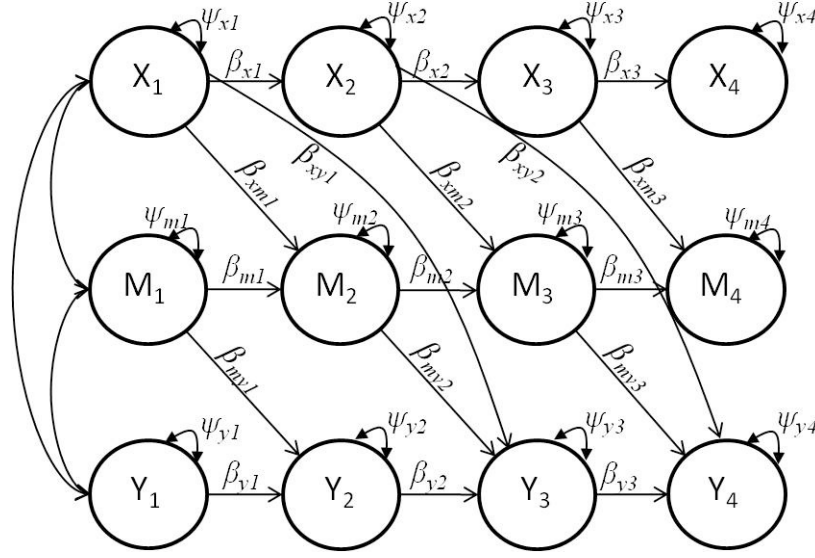


Figure 2.1: Cross-lagged Panel Model

The CLPM with four occasions of measurement for each of three variables: the predictor (X), the mediator (M), and the outcome (Y).

2.3 Multilevel Structural Equation Modeling

2.3.1 Foundations and Development

Despite the separate treatment of MLM and SEM in the literature and the different types of research questions for which the two techniques are each suited, the two analytical paradigms are largely similar. In fact, certain parameterizations can turn an MLM into an SEM. The relationship between MLM and SEM has been explored in recent literature (Bauer, 2003; Bovaird, 2007; Curran, 2003; Mehta & Neale, 2005).

Studies show how a MLM growth model can be respecified as a latent growth curve model in SEM, and argues that whereas the estimation procedures for the two techniques differ, the results are analytically identical (Curran, 2003; MacCallum et al., 1997). Mehta and Neale (2005) similarly demonstrate how a standard mixed effects multilevel model can be implemented in SEM, further arguing that the primary distinction between these two model specification methods is that in SEM the likelihood of the model parameters is specified at the level of clustering whereas in MLM it is specified for the entire data vector, which is essentially only a representational differ-

ence. Likewise, both Bauer (2003) and Mehta and Neale (2005) illustrate that the SEM measurement model can be fit as a MLM. Comparing the two models, it can be shown that, whereas the measurement model can be used to partition the observed variance into shared factor variance and unique residual variance, the MLM specification can be used to partition the variance into level-2 (between-group) and level-1 (within-group) variance (Mehta & Neale, 2005). So, for example, in the case of longitudinal data, the MLM level-1 variance between repeated measures is analogous to the SEM unique or residual variance, and the MLM level-2 variance between individuals is analogous to the shared factor variance in SEM.

Multilevel models can be reparameterized into structural equation models by specifying random effects as latent variables (Bovaird, 2007; Mehta & Neale, 2005). By doing this, many of the limitations of MLM can be overcome. Reparameterizing allows for the use of latent variables to correct for measurement error, multivariate outcomes, flexible multiple group comparisons, and the calculation of overall fit statistics for model evaluation (Bovaird, 2007; Curran, 2003; Mehta & Neale, 2005). In fact, any two-level MLM can be specified as a SEM, but increasingly complex models become difficult to specify and to estimate (Curran, 2003).

The two methods translate from one to the other easily enough that Curran (2003) examines where the two approaches differ and under what circumstances each method should be implemented. He concluded that the benefits of the SEM approach to modeling multilevel data suggest it should be used where possible, and the primary problem with using SEM is simply a data management issue; however, recent advances in estimation techniques make even data management less of a barrier to implementing these models.

2.3.2 Goldstein and McDonald (1986, 1988)

Goldstein and McDonald (1988) proposed a multilevel model for latent variables. Using a simple example of a two-level model of students in schools with an intercept, a level-1 predictor, and a level-2 predictor, the authors begin from a simple MLM given as:

$$y_{ij} = X_{0ij}\beta_{0j} + X_{1ij}\beta_{1ij} + X_{2ij}\beta_{2j}, \quad (2.14)$$

where, for an individual i in cluster j , β_1 represents the level-1, between-student-within-school vector with a level-1 covariance matrix Ω_1 , and β_2 represents the level-2, between-school vector with a level-2 covariance matrix Ω_2 .

The authors then define the latent model as a 2-level common factor model with the following equations:

$$\beta_{2j} = Aw_j + u_j, \quad (2.15)$$

and

$$\beta_{1ij} = Bz_{ij} + e_{ij}, \quad (2.16)$$

Then, by substituting 2.15 and 2.16 into 2.14 and re-arranging the equation to group coefficients and error terms, the multilevel factor model is given as:

$$y_{ij} = X_{0ij}\beta_{0j} + X_{1ij}Bz_{ij} + X_{2j}Aw_j + X_{1ij}e_{ij} + X_{2j}u_j. \quad (2.17)$$

The Goldstein and McDonald (1988) model presented here allows different levels of response variables to be modeled, as well as the flexibility to handle missing or unbalanced data. To begin, y_{ij} is decomposed into uncorrelated parts, y_{2j} and y_{1ij} , which represent the independent random sampling processes at level-2 and level-1. Then separate path models are fit for y_{2j} and y_{1ij} (McDonald, 1994). An iterative generalized least squares (IGLS) method of estimation is implemented

to provide efficient estimation in the presence of missing data. IGLS is outlined, and shown to be equivalent to full information maximum likelihood (FIML), in Goldstein (1986). IGLS allows estimation from raw data, rather than structured mean and covariance matrices, by iteratively estimating a covariance matrix, β_k , then using β_k to estimate the model parameters, then re-estimating β_k using the model parameters, and continuing to the next iteration. This estimation procedure gives two distinct advantages to the Goldstein and McDonald (1988) approach: the flexibility to handle unbalanced data and the ability to model random slope parameters. However, it is more computationally intensive than other methods.

2.3.3 Muthén's (1989) MUML Method

In contrast to Goldstein and McDonald, Muthén (1989) proposed a model for multilevel covariance structure analysis which uses covariance matrices rather than raw data as input and utilizes an approximate maximum likelihood method of estimation: MUML (Muthén's Maximum Likelihood). Like Goldstein and McDonald, Muthén's method breaks the model into between- and within-components, but in this case it involves decomposing the covariance matrix rather than y . Muthén begins his method with the factor analysis model:

$$y_{ij} = \nu + \Lambda\eta_{ij} + \varepsilon_{ij}, \quad (2.18)$$

where i represent an individual, j represents a cluster, $E(\varepsilon_{ij}) = 0$ and $V(\varepsilon_{ij}) = \Theta$. Then the level-1 and level-2 equations for the model are given as:

$$\eta_{ij} = \alpha_j + \omega_{ij}, \quad (2.19)$$

where ω_{ij} is a within-group-between-individual random component, α_j is a between-group random component further specified as:

$$\alpha_j = \alpha + \Gamma z_j + \delta_{\alpha_j}, \quad (2.20)$$

and z_j is a vector of group-level variables. The full model can then be obtained from 2.18, 2.19, and 2.20 as:

$$y_{ij} = \nu + \Lambda(\alpha + \Gamma z_j + \delta_{\alpha_j}) + \Lambda \omega_{ij} + \varepsilon_{ij}. \quad (2.21)$$

The decomposed covariance matrix, then, is:

$$V(y) = \Sigma_W + \Sigma_B, \quad (2.22)$$

with

$$\Sigma_W = \Lambda V(\omega) \Lambda' + \Theta \quad (2.23)$$

and

$$\Sigma_B = \Lambda \Gamma V(z) \Gamma' \Lambda' + \Lambda V(\delta_\alpha) \Lambda'. \quad (2.24)$$

By separating the covariance matrix into between and within components, the model can essentially be viewed as a two-group model where factor models are simultaneously fit to both the between- and the within-group covariance matrices as if each matrix represented a separate, independent sample. The parameters of the factor model can then be specified to vary or fixed to equality across groups (Muthén, 1989, 1994). Using this model, factor means may vary and be ex-

panded to allow for between-group variability of intercepts; however, in contrast to the Goldstein and McDonald estimation method, the model does not allow for estimation of random slope parameters. Also, the MUML estimator is a limited information maximum likelihood estimator and is not unbiased in cases of unbalanced data with regard to the between-group covariance matrix. To correct this problem a scaling parameter, c , is used,

$$c = [N^2 - \sum N_g^2] [N(G - 1)]^{-1}. \quad (2.25)$$

where G is the number of groups, N is the total number of observations, and N_g is the number of observations in each group. This scalar is then multiplied to Σ_B . When the group sizes are equal, c is simply the group size; with unbalanced data and a large number of groups, c is approximately the mean group size (Muthén, 1994). McDonald (1994) calls this method of weighting a covariance matrix from unbalanced data by class size “pseudobalanced.” The MUML estimation method allows for much easier computation than the Goldstein and McDonald IGLS method; however, MUML’s limitation with regard to random slope estimation and unbalanced data, along with modern computing capabilities, make the Goldstein and McDonald method seemingly preferable.

2.3.4 Current Models and Estimation

While the early Muthén model shows a simple case with a two-level factor model, recent extensions demonstrate the flexibility of these models and overcome some of the initial limitations of the method. Muthén and Asparouhov (2008) give an MSEM approach to growth mixture modeling which includes not only a simple measurement model, but also level-1 and level-2 structural models, which include vectors of random parameters and allow for the estimation of random slopes. Here, the initial measurement model is defined as:

$$Y_{ij} = \mathbf{v}_j + \Lambda_j \boldsymbol{\eta}_{ij} + K_j X_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad (2.26)$$

where i represents individuals at level-1, j represents clusters at level-2, v_j is a vector of intercepts, Λ_j is a slope matrix for η_{ij} , which is the vector of random effects, K_j is the slope matrix for the covariates, X_{ij} , and ε_{ij} is a vector of residuals, which are assumed to be multivariate normally distributed with a mean of zero and a covariance matrix, θ . The model allows for random effects as variables and elements of parameter matrices, v_j , Λ_j , and K_j can vary across cluster, j . The level-1 structural model is given as:

$$\eta_{ij} = \alpha_j + B_j \eta_{ij} + \Gamma_j X_{ij} + \zeta_{ij}, \quad (2.27)$$

where α_j is a vector of intercepts, B_j is a matrix of regression parameters, Γ_j is a slope matrix, and ζ_{ij} is a vector of residuals, which are assumed to be multivariate normally distributed with a mean of zero and a covariance matrix, Ψ . As before, elements of the parameter matrices, α_j , B_j , and Γ_j , can vary across cluster. The general model can be extended to a multilevel model with the level-2 structural model:

$$\eta_j = \mu + \beta \eta_j + \gamma X_j + \zeta_j. \quad (2.28)$$

In the level-2 structural model η_j is a vector of random effects, which includes the r random elements of each of the parameter matrices, v_j , Λ_j , K_j , α_j , B_j , and Γ_j . The estimated fixed effects are contained in μ , β , and γ . The vector, μ , is an $r \times 1$ vector of means of random effects and intercepts of the between-level equations. The β matrix is an $r \times r$ matrix of slopes of random effects on each other. The γ matrix is an $r \times s$ matrix of slope parameters of random effects on cluster-level variables.

Estimation of these models can be done using a maximum likelihood estimator with an EM algorithm, or through a number of alternative algorithms, including an accelerated EM algorithm

(AEM), which is the method currently implemented in Mplus. The EM algorithm involves estimating posterior distributions of latent variables and maximizing the log-likelihood function with respect to the model parameters. AEM estimation directly optimizes the maximum likelihood estimates by using standard maximization algorithms such as Fisher scoring or quasi-Newton algorithms (Muthén & Asparouhov, 2008). By changing the estimation method and including random effects vectors in the model, the new Muthén and Asparouhov model is able to overcome the limitations of the early Muthén model. Figure 3 is a graphical illustration of the general MSEM which includes equations for the measurement model as well as the between and within structural models. One random effect, the slope B_{MXj} , is included. This random effect is represented at the within level by a dark circle on the path, B_{MXj} , and at the between level by a latent variable.

2.3.5 Other issues related to MSEM

Sample size considerations in MSEM, as with all complex modeling techniques, are both important and complicated. Because of the multilevel structure, more than one sample size must be taken into account. In cases where the available number of groups is limited, determining and obtaining an appropriate sample size can become increasingly difficult. In an application of MSEM to cross-cultural research, Cheung and Au (2005) found that MSEM can be used effectively even when the group-level sample size is small, but with the caveat that increasing the sample size at the individual level does not compensate for the smaller sample size at the group level, further emphasizing that sample sizes at level 1 and level 2 have differential impacts on model estimation. MSEM has also been used to test measurement invariance in cross-cultural research, and under certain conditions (i.e., with a large population at the group-level from which to draw a large random sample), MSEM has advantages over manifest variable approaches and is potentially useful beyond current multiple group latent variable models for testing invariance, depending upon the research question of interest (Selig et al., 2008).

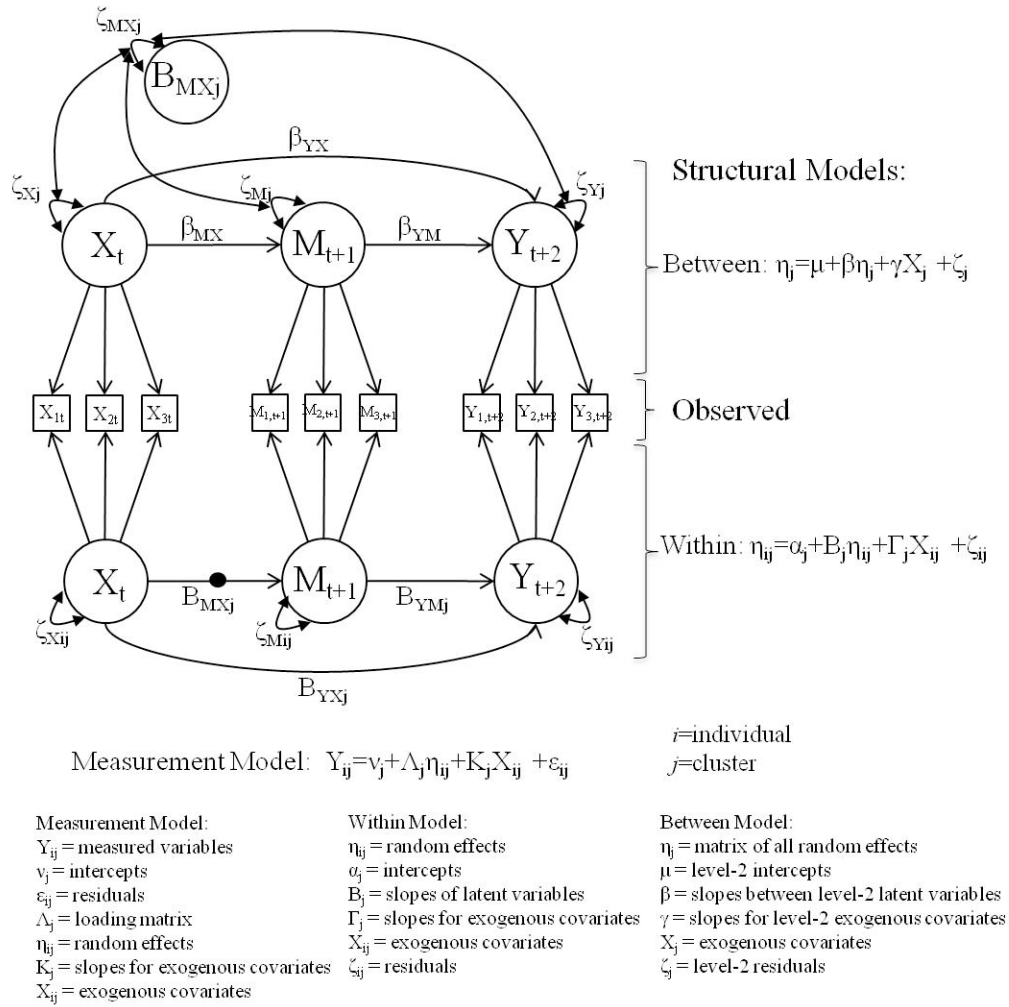


Figure 2.2: An example of an MSEM model.

Solid black circles on latent variable paths at the within represent an estimated random effect which is represented by an additional latent variable at the between level.

Like sample size considerations, the assessment of model fit in MSEM is not straightforward. Although MSEM does provide measures of overall model fit, the interpretation of these measures is not as clear as in single-level latent variable models. Again, the effect of having multiple levels contained within the model means that fit at all levels has to be considered. Therefore, a poorly fitting model may be the result of poor fit at the within- or the between-level or at both levels. Current solutions for this problem are to estimate each level of the model separately and obtain fit indices for each part of the model (Yuan & Bentler, 2007) or fitting a partially saturated model and comparing fit between models (Ryu & West, 2009).

Another consideration of these models is how to identify them. Having a factor that is endogenous at two levels means that fixing the scale in an appropriate and interpretable manner can present a challenge. Typical methods of model identification, such as fixing residual variances of the factor to one or fixing a loading to one, are not recommended. Instead, McDonald (1994) suggests fixing factor loadings to be equal across levels creating parallel forms, thus implying that the factor is the same at both levels, but with distinct residual variances. This leads to easier interpretation and smaller standard errors compared to other methods. Despite these considerations, the practical implementation of MSEM methods is well within reach of most researchers, and is quickly becoming more prevalent in the fields of psychology, sociology, and education.

Models in MSEM can be fit in two different ways. The first is as a reparameterization of the MLM where random effects are estimated as latent variables. The second way is to fit complete models at both the between- and within-level. In estimating an MSEM as a reparameterized MLM the model is specified as a SEM, but random effects are included as additional latent variables. The unit of analysis shifts from individual to cluster, and the model-implied means and covariances are the random intercepts (Mehta & Neale, 2005). Estimating an MSEM as a full between and within model involves first decomposing the covariance matrix into between and within components. Then separate models are fit to each covariance matrix. The restrictions placed on the between and within models can be different in both the measurement and structural models (Mehta & Neale, 2005). For example, the between model can have a different number of latent variables or different

paths among the latent variables than does the within model.

2.3.6 Mediation in MSEM

By moving mediation analysis into an MSEM framework, the problems of specifying these models in MLM can be overcome (Preacher et al., 2010). Using the model developed by Muthén and Asparouhov (2008), Preacher et al. (2010) propose a model for mediation in MSEM. The authors start by simplifying 2.26, 2.27, and 2.28 into

$$Y_{ij} = \Lambda \eta_{ij}, \quad (2.29)$$

$$\eta_{ij} = \alpha_j + B_j \eta_{ij} + \zeta_{ij}, \quad (2.30)$$

and

$$\eta_j = \mu + \beta \eta_j + \zeta_j, \quad (2.31)$$

respectively. From here, the authors show that for a single-level design (1-1-1) the general form for the MSEM model is:

$$Y_{ij} = (I - B)^{-1} \zeta_{ij} \quad (2.32)$$

where B is a matrix containing the path coefficients B_{MXj} and B_{YMj} , which make up the indirect effect. Calculating the total indirect effect using Bollen (1987), the effect of X on Y through M is $B_{MXj}B_{YMj}$. This same procedure can be used for models with upper level mediation such as 2-1-1

or 2-2-1 models. Also, because the effect is calculated as a product, the asymmetric confidence limit approach can be used to test the effect. Preacher et al. (2010) demonstrate the equations for a general MSEM model for a 2-1-1 mediation model as follows:

$$Y_{ij} = \begin{bmatrix} X_{ij} \\ M_{ij} \\ Y_{ij} \end{bmatrix} = \Lambda \eta_{ij} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} \quad (2.33)$$

$$\eta_{ij} = \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ B_{YMj} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{Mij} \\ \eta_{Yij} \\ \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{Mij} \\ \zeta_{Yij} \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (2.34)$$

$$\eta_j = \begin{bmatrix} B_{YMj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} = \begin{bmatrix} \mu_{BYMj} \\ \mu_{\alpha \eta Xj} \\ \mu_{\alpha \eta Mj} \\ \mu_{\alpha \eta Yj} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta_{MX} & 0 & 0 \\ 0 & \beta_{YX} & \beta_{YM} & 0 \end{bmatrix} \begin{bmatrix} B_{YMj} \\ \alpha_{\eta Xj} \\ \alpha_{\eta Mj} \\ \alpha_{\eta Yj} \end{bmatrix} + \begin{bmatrix} \zeta_{BYMj} \\ \zeta_{\alpha \eta Xj} \\ \zeta_{\alpha \eta Mj} \\ \zeta_{\alpha \eta Yj} \end{bmatrix} \quad (2.35)$$

Then, using Bollen's (1987) formula, the between indirect effect is given as:

$$F = (I - \beta_B)^{-1} - I - \beta_B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{MX}\beta_{YM} & 0 & 0 \end{bmatrix} \quad (2.36)$$

The MSEM model has been compared to MLM and an unconfated multilevel model (UMM;

Zhang et al., 2009) to compare the effectiveness of each method in terms of bias, confidence interval estimation, efficiency, convergence, and power. A simulation study found that for models that include upper-level mediation, MSEM reduces bias, improves confidence interval coverage, and is efficient in cases with larger ICC and cluster sizes (Preacher, Zhang, & Zyphur, 2011). Testing mediation in this way not only compensates for the shortcomings of MLM, but also adds distinct advantages. Random slopes can not only be included in the model, they can be used as mediator or predictor variables. MSEM also allows for the assessment of overall model fit, which is not possible in MLM because of the lack of a saturated model to use as a comparison (Preacher et al., 2010).

Using MSEM for mediation combines the advantages of MLM for handling clustered data and the advantages of SEM for modeling different types of heterogeneity across time. MSEM also allows for level-2 outcome variables where MLM does not. Whereas the potential advantages of MSEM for mediation are numerous, models for longitudinal mediation have not been examined in this framework.

2.3.7 Current Study

The cross-lagged panel model proposed by Cole and Maxwell (2003) appropriately handles mediation for longitudinal data, but it is limited in that it can be used to assess only interindividual change, not intraindividual change or individual differences in change (Selig & Preacher, 2009). These heterogeneous effects are often the most interesting for developmental researchers. Therefore, methods that have the flexibility to include different sources of heterogeneity are of great interest. In the context of mediation models, this includes the ability to model heterogeneity in indirect effects.

The importance of longitudinal data for mediation, the advantages of MSEM for mediation, and the lack of longitudinal MSEM mediation models are the motivation for the current study. The primary aim of the study is to extend the work of Preacher et al. (2010) with applications of mediation in MSEM to longitudinal data and modeling both between between and within group sources

of heterogeneity in longitudinal data by using a multilevel cross-lagged panel model. Previous research has demonstrated the effect of ignoring the multilevel structure of data in path models as well as in a few common models in SEM such as the growth mixture model and MIMIC models. A similar approach will be used to examine the consequences of ignoring nesting in the CLPM, with a particular focus on bias in the indirect effect.

Given prior research into the effects of ignoring clustering in multilevel data, it is hypothesized that when level-2 heterogeneity is ignored, the estimates of the indirect effect and the associated standard errors will be biased and underestimated, and power for detecting the effects will be lower. Additionally, it is expected that the magnitude of the effects will increase as the ICC increases and as sample size decreases. Finally, the model fit for the MSEM should be better than the model fit of the single level model as it is a more appropriate model for the nested structure of the data.

Chapter 3

Methods

3.1 Simulation Parameters

A simulation study is used to demonstrate the consequences of ignoring heterogeneity in longitudinal models. Data were generated from a two-level CLPM, with three occasions of measurement for the outcome, two occasions of measurement for the mediator, and a single occasion of measurement for the predictor (see Figure 4). The path coefficients at level one were determined based on an estimated CLPM established in prior research (Maxwell & Cole, 2007). The path coefficients for level two were then altered to create discrepancy in the relationships among the latent variables across levels. Each level-2 path coefficient is of a smaller magnitude than its level-1 counterpart.

The values of the intraclass correlation (ICC) were set at .1, .25, and .4, and all variables within an ICC condition were given the same ICC. These ICCs are within the range sufficient for justifying MSEM as suggested by prior research (Cheung & Au, 2005; Heck & Thomas, 2009). The multilevel structure of the generating model means that two factors have to be considered for varying conditions of sample size: the number of clusters and cluster size. The simulation includes level-2 sample sizes across three conditions (50, 100, 250) and balanced level-1 sample sizes across two conditions (10 and 20). Previous studies using multilevel MIMIC models (Finch & French,

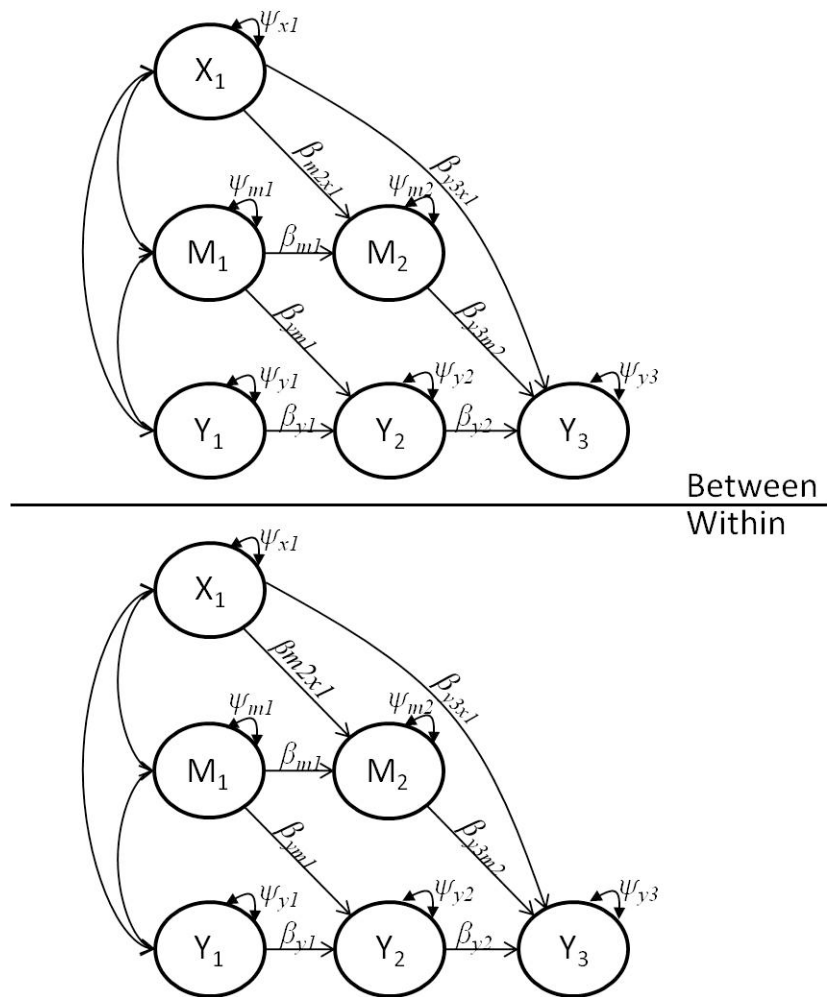


Figure 3.1: MSEM Simulation Data Generating Model

2011) and multilevel CFA models (Maas & Hox, 2005) have suggested these as appropriate sample sizes for MSEM.

Using Mplus software, 5000 replications of data are generated for each condition. Both the appropriate two-level and a single-level model (see Figure 5) are then fit to the data to examine bias, confidence interval coverage, and Type I error rates for the indirect effect. Bias is assessed by comparing the estimated indirect effects from the single level model to the known indirect effect in the population as well as the estimates from the two level model.

3.1.1 Determining Bias

The estimates from both the single- and two-level models were compared to population parameters to determine bias. Bias in standard error was also examined using estimated standard errors in comparison to standard deviations of estimates across replications. Relative bias in the indirect effect was calculated following the recommendations of Chen et al. (2010) and is as follows:

$$B(\bar{\theta}) = \frac{\bar{\theta}_{est} - \theta_{pop}}{\theta_{pop}} \quad (3.1)$$

where $\bar{\theta}$ is the mean of the estimate of the indirect effect across replications and θ_{pop} is the population parameter. For relative bias in parameter estimates, the acceptable level of bias is determined using the cutoff of 0.05 suggested by Hoogland and Boomsma (1998). Likewise, bias in the standard error of the indirect effect is calculated as follows:

$$B(\hat{S}_{\theta}) = \frac{\bar{\hat{S}}_{\hat{\theta}false} - S_{\hat{\theta}true}}{S_{\hat{\theta}true}} \quad (3.2)$$

where $\bar{\hat{S}}_{\hat{\theta}false}$ is the mean of the estimated standard error of the indirect effect across replications and $S_{\hat{\theta}true}$ is the standard deviation of the parameter estimate across replications. The acceptable level of relative bias in estimated standard errors is determined using the cutoff of 0.10 (Hoogland & Boomsma, 1998).

3.1.2 Confidence Interval Coverage and Power

In addition to examining bias, the MSEM and single level models are compared for confidence interval coverage and power. Confidence interval coverage is calculated from the proportion of replications for which the 95% confidence interval for a given effect included the true, population value of the indirect effect. Likewise, power was determined as the proportion of replications for which the null hypothesis that the estimated parameter is equal to the population value of the parameter was rejected.

3.1.3 Model Fit

Finally, the appropriateness of the models will be evaluated by examining model fit. Two model fit statistics are examined: the root mean square error of approximation (RMSEA; Browne & Cudeck, 1993) and the standardized root mean square residual (SRMR). The RMSEA, a commonly used test of model fit, is a measure of discrepancy between the observed and reproduced covariance matrices (Ryu & West, 2009). RMSEA is included as measure of model fit because of the common use and widespread understanding of the measure. SRMR is included as a measure of fit because for the MSEM, SRMR is calculated at both the between and within levels. Model fit in MSEM can be difficult to interpret because any misfit could be the result of either the between or the within level model (Ryu & West, 2009). Although, level-specific measure of SRMR may not accurately reflect true model fit at each level, it is used here to provide more information about the model at each level.

3.2 Results

3.2.1 Bias

Bias in the estimates of the indirect effect in the MSEM was well below the 0.05 cutoff for acceptable in both the within and between portions of the MSEM (see Table 1). For the single level

Table 3.1: Bias in the Indirect Effect

ICC	Cluster	n	MSEM Estimate Between		MSEM Estimate Within		Single Level Estimate		
			Estimate (PV = 0.0756)	Relative Bias	Estimate (PV = 0.0224)	Relative Bias	Estimate	PV	Relative Bias
.1	50	10	0.0759	0.004	0.0235	0.049	0.0202	0.0263	-0.232*
		20	0.0762	0.008	0.0227	0.013	0.0202	0.0263	-0.232*
	100	10	0.0761	0.007	0.0226	0.009	0.0203	0.0263	-0.228*
		20	0.0763	0.009	0.0224	0.000	0.0203	0.0263	-0.228*
	250	10	0.0765	0.012	0.0224	0.000	0.0204	0.0263	-0.224*
		20	0.0760	0.005	0.0224	0.000	0.0204	0.0263	-0.224*
.25	50	10	0.0758	0.003	0.0225	0.004	0.0238	0.0328	-0.274*
		20	0.0760	0.005	0.0225	0.004	0.0238	0.0328	-0.274*
	100	10	0.0768	0.016	0.0225	0.004	0.0241	0.0328	-0.265*
		20	0.0757	0.001	0.0225	0.004	0.0240	0.0328	-0.268*
	250	10	0.0764	0.011	0.0224	0.000	0.0200	0.0328	-0.390*
		20	0.0759	0.004	0.0224	0.000	0.0241	0.0328	-0.265*
.4	50	10	0.0759	0.004	0.0224	0.000	0.0270	0.0399	-0.323*
		20	0.0755	-0.001	0.0225	0.004	0.0268	0.0399	-0.328*
	100	10	0.0761	0.007	0.0225	0.004	0.0274	0.0399	-0.313*
		20	0.0762	0.008	0.0224	0.000	0.0272	0.0399	-0.318*
	250	10	0.0764	0.011	0.0224	0.000	0.0276	0.0399	-0.308*
		20	0.0762	0.008	0.0224	0.000	0.0274	0.0399	-0.313*

PV = Population value

* Value exceeds 0.05 cutoff, indicating an unacceptable level of bias

model, the population value for determining relative bias was derived from a weighted average of the values of the MSEM between and within indirect effects based on ICC. An acceptable level of bias was exceeded in every sample size and ICC condition in the single level model.

With only one exception, bias in the estimated standard error of the indirect effect was at or below acceptable levels across all conditions for both the between and within level models of the MSEM (see Table 2). The 0.10 criterion for an acceptable level of bias was exceeded in the single level model in almost every condition. The estimated standard error of the indirect effect decreased as sample size increased for both levels of the MSEM as well as for the single level model. For the single level model, the amount of bias increased as the sample size and ICC increased.

Table 3.2: Bias in the Estimated Standard Error of the Indirect Effect

ICC	Cluster	n	MSEM Estimate Between			MSEM Estimate Within			Single Level Estimate		
			Standard Error	Standard Deviation	Relative Bias	Standard Error	Standard Deviation	Relative Bias	Standard Error	Standard Deviation	Relative Bias
.1	50	10	0.0782	0.0754	0.037	0.0103	0.0099	0.040	0.0109	0.0113	-0.035
		20	0.0829	0.0824	0.006	0.0072	0.0068	0.059	0.0076	0.0083	-0.084
	100	10	0.0640	0.0623	0.027	0.0073	0.0069	0.058	0.0076	0.0080	-0.050
		20	0.0577	0.0576	0.002	0.0052	0.0047	0.106*	0.0054	0.0060	-0.100*
	250	10	0.0437	0.0435	0.005	0.0045	0.0043	0.047	0.0048	0.0051	-0.059
		20	0.0335	0.0333	0.006	0.0028	0.0029	-0.034	0.0034	0.0038	-0.105*
.25	50	10	0.0761	0.0755	0.008	0.0105	0.0099	0.061	0.0107	0.0135	-0.207*
		20	0.0610	0.0604	0.010	0.0071	0.0067	0.060	0.0075	0.0111	-0.324*
	100	10	0.0501	0.0497	0.008	0.0072	0.0068	0.059	0.0075	0.0095	-0.211*
		20	0.0409	0.0397	0.030	0.0051	0.0047	0.085	0.0053	0.0081	-0.346*
	250	10	0.0300	0.0293	0.024	0.0044	0.0043	0.023	0.0047	0.0060	-0.217*
		20	0.0243	0.0242	0.004	0.0030	0.0030	0.000	0.0033	0.0051	-0.353*
.4	50	10	0.0727	0.0712	0.021	0.0101	0.0098	0.031	0.0106	0.0169	-0.373*
		20	0.0545	0.0531	0.026	0.0071	0.0067	0.060	0.0074	0.0149	-0.503*
	100	10	0.0411	0.0408	0.007	0.0069	0.0068	0.015	0.0074	0.0118	-0.373*
		20	0.0420	0.0408	0.029	0.0045	0.0045	0.000	0.0052	0.0108	-0.519*
	250	10	0.0436	0.0429	0.016	0.0046	0.0043	0.070	0.0046	0.0073	-0.370*
		20	0.0442	0.0440	0.005	0.0029	0.0028	0.036	0.0032	0.0068	-0.529*

*Value exceeds 0.10 cutoff, indicating an unacceptable level of bias

3.2.2 Confidence Interval Coverage and Power

The 95% confidence interval coverage for the indirect effect reached the 95% most consistently in the within level of the model (see Table 3). Although the confidence interval coverage did not reach the 95% level in most of the conditions for the between level of the MSEM, the accuracy was worse for the single level model, where even the conditions with the most coverage never reached 95%. Accuracy increased with sample size in both levels of the MSEM, but decreased with sample size in the single level model.

Power for the single level model was good in many conditions, reaching 1.0 in the higher sample size conditions, but was low in small sample size conditions, especially when the ICC was high (see Table 3). The between level of the MSEM had low power across all conditions, reaching above 0.8 only in the highest sample size conditions. Contrary to expectations, the power for the single-level model when clustering was ignored was not substantially lower than in the MSEM.

Table 3.3: Coverage and Power

ICC	Cluster	n	95% CI Coverage Rate			Power		
			Between	Within	Single	Between	Within	Single
.1	50	10	0.963	0.963	0.842	0.000	0.363	0.591
		20	0.918	0.941	0.787	0.012	0.882	0.933
	100	10	0.929	0.966	0.783	0.013	0.890	0.936
		20	0.925	0.953	0.695	0.095	0.941	0.999
	250	10	0.939	0.973	0.667	0.105	0.944	1.000
		20	0.947	0.963	0.488	0.566	0.970	1.000
.25	50	10	0.911	0.924	0.644	0.041	0.593	0.902
		20	0.913	0.934	0.621	0.121	0.905	0.917
	100	10	0.921	0.946	0.692	0.205	0.923	0.931
		20	0.933	0.948	0.552	0.385	0.932	0.996
	250	10	0.950	0.959	0.231	0.737	0.915	1.000
		20	0.950	0.961	0.230	0.883	0.941	1.000
.4	50	10	0.900	0.923	0.817	0.123	0.621	0.540
		20	0.914	0.934	0.521	0.189	0.923	0.832
	100	10	0.920	0.943	0.576	0.398	0.923	0.894
		20	0.969	0.983	0.431	0.455	0.951	0.976
	250	10	0.950	0.989	0.385	0.737	0.955	0.999
		20	0.990	0.990	0.223	0.878	0.991	1.000

However, as ICC increased the power was higher for the within level of the MSEM than for the single level model in smaller sample size conditions.

3.2.3 Model Fit

Model fit was examined using mean values of the RMSEA over the 5,000 replications. Across all conditions, model fit was better in the MSEM than in the single level model (see Table 4). Whereas the RMSEA indicates good or close for all sample size and ICC conditions of the MSEM, the model fit for the single-level model based on RMSEA was acceptable in almost all of the sample size conditions for ICC = 0.1 but not in the higher ICC conditions.

The model fit based on SRMR showed similar patterns. Although the fit was less good for the single level model than for either level of the MSEM, it was still within acceptable limits. The SRMR also showed better model fit at the within level than the between level of the MSEM.

Table 3.4: Model Fit

ICC	Cluster	N	RMSEA MSEM	RMSEA Single	SRMR Between	SRMR Within	SRMR Single
.1	50	10	0.020**	0.110	0.082	0.020^	0.063^
		20	0.012**	0.096*	0.066^	0.014^	0.054^
	100	10	0.011**	0.096*	0.063^	0.015^	0.053^
		20	0.007**	0.096*	0.047^	0.010^	0.052^
	250	10	0.006**	0.096*	0.041^	0.009^	0.051^
		20	0.004**	0.096*	0.030^	0.006^	0.050^
.25	50	10	0.016**	0.110	0.059^	0.020^	0.063^
		20	0.011**	0.110	0.051^	0.014^	0.061^
	100	10	0.009**	0.109	0.041^	0.014^	0.059^
		20	0.006**	0.109	0.035^	0.010^	0.058^
	250	10	0.006**	0.095*	0.026^	0.009^	0.051^
		20	0.004**	0.109	0.022^	0.006^	0.056^
.4	50	10	0.015**	0.126	0.050^	0.020^	0.073^
		20	0.011**	0.127	0.047^	0.014^	0.069^
	100	10	0.009**	0.124	0.036^	0.014^	0.065^
		20	0.006**	0.124	0.031^	0.010^	0.064^
	250	10	0.006**	0.122	0.026^	0.009^	0.062^
		20	0.004**	0.122	0.020^	0.006^	0.061^

*RMSEA value indicates acceptable fit

**RMSEA value indicates good or close fit

^SRMR value indicated good fit

3.3 Discussion

As expected, the most bias was in the estimates and estimated standard errors of the indirect effect in the single level model, supporting the first hypothesis of the study. Additionally, the negative direction of the bias indicates that the standard errors were underestimated with regard to the standard deviation across replications, which was also expected.

Also as expected, the amount of bias increases as the ICC increases under most conditions. For the single level model, however, the amount of bias in the estimated standard errors is often larger in the larger cluster size conditions, which is contrary to the expected result that bias would decrease as sample size increases.

The coverage rate for both the between and within levels of the MSEM model was high, meeting or exceeding 95% coverage in many conditions of the within model and in large sample size conditions for the between model. The single level model never reached 95% coverage, and the accuracy of the single level model was lowest in the higher sample size conditions, particularly

when ICC was high. The low accuracy in high ICC condition is expected given that with clustered data, low ICC conditions are more similar single level data than multilevel data with a high ICC.

The power for between and within level as well as single level models only partially supported the hypothesis and prior research. While the single level had power of 1.0 in higher sample size conditions, neither the between nor the within level of the MSEM had power of 0.8 or higher in most conditions. This finding is in contrast to that of prior research showing that power is reduced when a level is ignored (Moerbeek, 2004). Also, the power for the between level was low, and did not reach acceptable levels in most conditions. Preacher et al. (2011) also found low power for the between level indirect effect when the number of clusters was less than 300. The largest sample size in the current simulation is 250 clusters and falls below this mark, suggesting that the low power in the between level effect may be the result of small sample size.

The multilevel structure of the data suggests that an MSEM should fit the data better than a single level model, and this hypothesis was supported. It should be noted, however, that strictly using SRMR as a criterion, the single level model still fit the data reasonably well, particularly in the low ICC conditions. It is not surprising that the single level model has better fit in low ICC conditions as there is proportionally less between-level variation to be ignored than at higher levels of ICC, and as a result, the consequences of ignoring the clustering are less. Finch and French (2011) also found acceptable fit when applying an inappropriate, single-level model to multilevel data while still observing unacceptable Type I error rates. This finding demonstrates that a model with good fit is not necessarily a good model. Given the level of bias observed in the single level model estimates and estimated standard errors of the indirect effect, accepting the single level model on the basis of model fit alone could lead to spurious conclusions.

3.4 Future Directions

The current study examined only the case of balanced designs. Although previous studies of the effects of ignoring a level of clustered data have examined both balanced and unbalanced

designs, the extension of this work to longitudinal data, where missing occasions commonly cause unbalanced data, is of interest. Furthermore, although Van den Noortgate et al. (2005) found no differences in the outcomes for models with balanced versus unbalanced designs, Moerbeek (2004) found different results for balanced and unbalanced designs. These discrepant findings warrant further investigation.

Another area for future research is to address the problems with the between level of the MSEM. While the bias, coverage, and standard errors of the within level of the MSEM supported the hypotheses of the study, the between level showed some unexpected results with regard to power. Testing models with varying degrees and direction of discrepancy across the within and between levels of the model may provide better between level results. Differing levels of discrepancy may also affect the amount of bias found in the single level model results.

Finally, some basic extensions could be made. The low power, bias, and poor model fit could be examined through additional, larger sample size conditions. Also, power was examined as an evaluation of the appropriateness of the MSEM versus the single level model, but Type I error has not been addressed and could provide additional information about the consequences of ignoring nesting, with specific regard to the estimation of the indirect effect. Although the indirect effect was of primary interest for this study, additional parameters could also be examined for bias. Previous research has examined the differential consequences of ignoring nesting for fixed effect and variance components, and a similar question could be posed for the longitudinal MSEM.

3.5 Conclusion

Four hypotheses for the consequences of ignoring a level of heterogeneity in an MSEM were laid out with specific regard to the indirect effect. The first, that estimates of the indirect effect and its associated standard error would be biased in the single level model, was supported by the simulation study. However, the between level of the MSEM also showed a degree of bias in the estimation of the indirect effect, although the estimated standard error did not have the same bias.

The second hypothesis was also supported. The estimated standard errors in the single level model were underestimated.

The final two hypotheses, addressing power and model fit, were not supported. Whereas the within level of the MSEM performed as expected, the single-level model had higher power and better model fit than was hypothesized while the between level of the MSEM had lower power and worse model fit than expected. These two issues may be resolved with larger sample sizes in the simulation.

Although the single level model often had higher power in many conditions, the degree of bias in the parameter and standard error estimates suggests that use of the single level model could lead to inappropriate conclusions. The single level model, particularly as ICC increases, loses power to detect the indirect effect and the estimate of the indirect effect and its associated standard error are biased and underestimated. Additionally, the single level model is less accurate and more biased in estimating both the indirect effect and the standard error as sample size and ICC increase, conditions under which estimating the MSEM becomes more powerful, accurate, and efficient.

These findings have serious implications for practical research. First, if power is low, the process of mediation by estimating an indirect effect may not be discovered when it is actually present. Second, even if the indirect effect is detected, common interpretations using confidence intervals and effect sizes will be inaccurate because of the bias in the parameter estimate and estimated standard error. In general, MSEM should be preferred to a single level model, though accurate and efficient estimation of MSEM requires a much larger sample size than a comparable single level model.

References

- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.
- Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, 28(2), 135–167.
- Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology*, 17, 37–69.
- Bovaird, J. A. (2007). Multilevel structural equation models for contextual factors. In B. J. A. Little, T. D. & N. A. Card (Eds.), *Modeling Contextual Effects in Longitudinal Studies* (pp. 149–182). Mahwah, NJ: Lawrence Erlbaum Associates.
- Browne, M. W. & Cudeck, R. (1993). *Testing structural equation models*, chapter Alternative ways of assessing model fit, (pp. 136–162). Sage: Newbury Park, CA.
- Campbell, D. T. (1963). *Problems in the Measurement of Change*, chapter From description to experimentation: Interpreting trends as quasi-experiments, (pp. 212–254). University of Wisconsin Press, Madison.
- Chen, Q., Kwok, O., Luo, W., & Willson, V. (2010). The impact of ignoring a level of nesting structure in multilevel growth mixture models: A monte carlo study. *Structural Equation Modeling*, 17, 570–589.

- Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2003). Investigation of mediational process using parallel process latent growth curve modeling. *Structural Equation Modeling*, 10, 238–262.
- Cheung, G. W. & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, 11(2), 296–325.
- Cheung, M. & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling*, 12(4), 598–619.
- Cole, D. A. & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112(4), 558–577.
- Curran, P. J. (2003). Have multilevel models been structural models all along? *Multivariate Behavioral Research*, 38(4), 529–569.
- Finch, W. H. & French, B. F. (2011). Estimation of mimic model parameters with multilevel data. *Structural Equation Modeling*, 18, 229–252.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43–56.
- Goldstein, H. & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, 53(4), 455–467.
- Gollob, H. F. & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, 58, 80–92.
- Gollob, H. F. & Reichardt, C. S. (1991). *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*, chapter Interpreting and estimating indirect effects assuming time lags really matter, (pp. 243–259). Washington DC: American Psychological Association.

- Hayes, A. F. (2009). Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408–420.
- Heck, R. H. & Thomas, R. P. (2009). Multilevel structural equation models. In R. H. Heck & R. P. Thomas (Eds.), *An Introduction to Multilevel Modeling Techniques* (pp. 134–162). New York, NY: Routledge, 2nd edition.
- Hoffman, L. & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development*, 6(2), 97–120.
- Hoogland, J. J. & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods and Research*, 26, 329–367.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: User's reference guide*. Chicago, IL: Scientific Software International.
- Judd, C. M. & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602–619.
- Krull, J. L. & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23(4), 418–444.
- Krull, J. L. & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36(2), 249–277.
- Maas, C. J. M. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1, 86–92.
- MacCallum, R. C., Kim, C., Malarkey, W. B., & Kiecolt-Glaser, J. K. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, 32(3), 215–253.

- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7, 83–104.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99–128.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41–62.
- Maxwell, S. E. & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12(1), 23–44.
- McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods Research*, 22(3), 399–413.
- Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). *Selected tables in mathematical statistics: Vol. VII: The product of two normally distributed random variables*. Providence, RI: American Mathematical Society.
- Mehta, P. D. & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, 10(3), 259–284.
- Moerbeek, M. (2004). The consequences of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research*, 39(1), 129–149.
- Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54(4), 557–585.
- Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods Research*, 22(3), 376–398.

- Muthén, B. & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal Data Analysis*. Boca Raton: Chapman and Hall/CRC Press.
- Muthén, L. K. & Muthén, B. O. (1998-2010). *Mplus User's Guide*. Los Angeles, CA: Muthén and Muthén, sixth edition.
- Pitariu, A. H. & Ployhart, R. E. (2010). Explaining change: Theorizing and testing dynamic mediated longitudinal relationships. *Journal of Management*, 36(2), 405–429.
- Pituch, K. A., Whittaker, T. A., & Stapleton, L. M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research*, 40(1), 1–23.
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel sem. *Structural Equation Modeling*, 18, 161–182.
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel sem framework for assessing multilevel mediation. *Psychological Methods*, 15, 209–233.
- R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Application and data analysis methods*. London: Sage.
- Roberts, J. K. (2004). An introductory primer on multilevel and hierarchical linear modeling. *Learning Disabilities: A Contemporary Journal*, 2(1), 30–38.
- Ryu, E. & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, 16, 583–601.
- Selig, J. P., Card, N. A., & Little, T. D. (2008). Latent variable structural equation modeling in cross-cultural research: Multigroup and multilevel approaches.

- Selig, J. P. & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, 6(3), 144–164.
- Skinner, C. & de Toledo Vieira, M. (2007). Variance estimation in the analysis of clustered longitudinal survey data. *Survey Methodology*, 33(1), 3–12.
- Snijders, T. & Bosker, R. (1999). *Multilevel analysis*. London: Sage.
- Sobel, M. E. (1982). *Asymptotic confidence intervals for indirect effects in structural equation models*, (pp. 290–312). American Sociological Association: Washington, DC, sociological methodology 1982 edition.
- Van den Noortgate, W., Opdenakker, M., & Onghena, P. (2005). The effects of ignoring a level of nesting in multilevel analysis. *School Effectiveness and School Improvement*, 16(3), 281–303.
- Yuan, K. H. & Bentler, P. M. (2007). Multilevel covariance structure analysis by fitting multiple single-level models. *Sociological Methodology*, 37, 53–82.
- Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, 12(4), 695–719.